



**PATENT APPLICATION**

A

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Docket No: 29284/35302

**PATENT APPLICATION TRANSMITTAL UNDER 37 C.F.R. 1.53**

**Box Patent Application**  
**Assistant Commissioner for Patents**  
**Washington, D.C. 20231**



Sir:

Transmitted herewith for filing is the patent application of

Inventor(s): Mitchel Kriss and Peter Cibulskis

Title: METHOD AND SYSTEM FOR MARKET RESEARCH DATA MINING

**1. Type of Application**

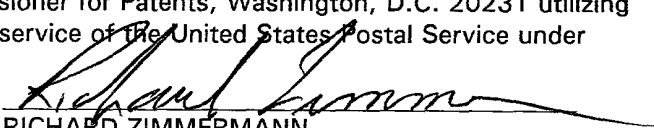
- ☒ This is a new application for a
- ☒ utility patent.
- ☐ design patent.
- ☐ This is a continuation-in-part application of prior application no.

**2. Application Papers Enclosed**

- |                                     |  |
|-------------------------------------|--|
| 1                                   | Title Page   |
| 28                                  | Pages of Specification (excluding Claims, Abstract & Drawings) |
| 14                                  | Page(s) of Claims  |
| 1                                   | Page(s) of Abstract  |
| 1                                   | Appendix   |
| 3                                   | Sheet(s) of Drawings (Figs. 1 to 2B)                           |
| <input checked="" type="checkbox"/> | Formal   |
| <input type="checkbox"/>            | Informal   |

**CERTIFICATION UNDER 37 CFR 1.10**

I hereby certify that this Patent Application Transmittal and the documents referred to as enclosed therewith are being deposited with the United States Postal Service on **June 10, 1999**, in an envelope addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231 utilizing the "Express Mail Post Office to Addressee" service of the United States Postal Service under Mailing Label No. EM099536707US.

  
RICHARD ZIMMERMANN

### 3. Declaration or Oath

- ☐ Enclosed
  - ☐ Executed by (check all applicable boxes)
    - ☐ Inventor(s)
    - ☐ Legal representative of inventor(s)  
(37 CFR 1.42 or 1.43)
    - ☐ Joint inventor or person showing a proprietary interest on behalf of inventor who refused to sign or cannot be reached
      - ☐ The petition required by 37 CFR 1.47 and the statement required by 37 CFR 1.47 are enclosed. See Item 5D below for fee.
- ☒ Not enclosed - the undersigned attorney or agent is authorized to file this application on behalf of the applicant(s). An executed declaration will follow.

### 4. Additional Papers Enclosed

- ☐ Preliminary Amendment
- ☐ Information Disclosure Statement
- ☐ Declaration of Biological Deposit
- ☐ Computer readable copy of sequence listing containing nucleotide and/or amino acid sequence
- ☐ Microfiche computer program
- ☐ Verified statement(s) claiming small entity status under 37 CFR 1.9 and 1.27
- ☐ Power of Attorney
- ☐ Verified translation of a non-English patent application
- ☐ An assignment of the invention
- ☒ Return receipt postcard
- ☐ Other

**5. Priority Applications Under 35 USC 119**

Certified copies of applications from which priority under 35 USC 119 is claimed are listed below and

☐ are attached.

☐ will follow.

COUNTRY	APPLICATION NO.	FILED

**6. Filing Fee Calculation (37 CFR 1.16)**

**A. ☒ Utility Application**

CLAIMS AS FILED - INCLUDING PRELIMINARY AMENDMENT (IF ANY)						
			SMALL ENTITY		OTHER THAN A SMALL ENTITY	
	NO. FILED	NO. EXTRA	RATE	FEE	RATE	FEE
BASIC FEE				\$380.00		\$760.00
TOTAL	49 - 20	= 29	X 9 =	\$	X 18 =	\$ 522.00
INDEP.	4 - 3	= 1	X 39 =	\$	X 78 =	\$ 78.00
<input type="checkbox"/> First Presentation of Multiple Dependent Claim			+ 130 =	\$	+ 260 =	\$ 0
Filing Fee:				\$	<b>OR</b>	\$1,360.00

**B. ☐ Design Application (\$155.00/\$310.00)** Filing Fee: \$ \_\_\_\_\_

**C. ☐ Plant Application (\$240.00/\$480.00)** Filing Fee: \$ \_\_\_\_\_

**D. Other Fees**

☐ Recording Assignment [Fee -- \$40.00 per assignment] \$ 0

☐ Petition fee for filing by other than all the inventors  
or person on behalf of the inventor where inventor refused  
to sign or cannot be reached [Fee -- \$130.00] \$ \_\_\_\_\_

☐ Other \$ \_\_\_\_\_

**Total Fees Enclosed \$1,360.00**

**7. Method of Payment of Fees**

- ☒ Enclosed is a check in the amount of: \$1,360.00
- ☐ Charge Deposit Account No. 13-2855 in the amount of: \$ \_\_\_\_\_  
A copy of this Transmittal is enclosed.
- ☐ Not enclosed

**8. Deposit Account and Refund Authorization**

The Commissioner is hereby authorized to charge any deficiency in the amount enclosed or any additional fees which may be required during the pendency of this application under 37 CFR 1.16 or 37 CFR 1.17 or under other applicable rules (except payment of issue fees), to Deposit Account No. 13-2855. A copy of this Transmittal is enclosed.

Please refund any overpayment to Marshall, O'Toole, Gerstein, Murray & Borun at the address below.

Please direct all future communications to Trevor B. Joike, at the address below.

Respectfully submitted,

MARSHALL, O'TOOLE, GERSTEIN,  
MURRAY & BORUN  
6300 Sears Tower  
233 South Wacker Drive  
Chicago, Illinois 60606-6402  
(312) 474-6300  
(312) 474-0448 (Telefacsimile)

By: \_\_\_\_\_

Trevor B. Joike  
Reg. No: 25,542

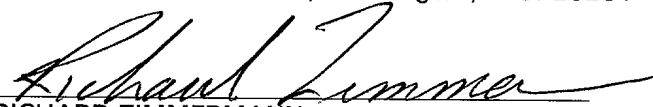
June 10, 1999

## JOINT INVENTORS

"EXPRESS MAIL" mailing label No. EM099536707US.

Date of Deposit: **June 10, 1999**

I hereby certify that this paper (or fee) is being deposited with the United States Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 CFR §1.10 on the date indicated above and is addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231

  
RICHARD ZIMMERMANN

## APPLICATION FOR UNITED STATES LETTERS PATENT

# S P E C I F I C A T I O N

### TO ALL WHOM IT MAY CONCERN:

Be it known that we, Mitchel Kriss, a citizen of the United States, residing at 1478 RFD, Long Grove, 60047, in the County of Lake and State of Illinois; and Peter Cibulskis, a citizen of the United States, residing at 600 S. Dearborn, Chicago, 60605, in the County of Cook and State of Illinois have invented a new and useful METHOD AND SYSTEM FOR MARKET RESEARCH DATA MINING, of which the following is a specification.

METHOD AND SYSTEM FOR MARKET RESEARCH DATA MINING

Technical Field of the Invention

The present invention is directed to a method and system for mining data, such as market research data.

Background of the Invention

Data are collected and stored in a database or a data warehouse for a variety of reasons. For example, it is known to collect market research data from a panel of product purchasers so that conclusions about the buying habits of specific population segments may be made. One such panel is operated by the A.C. Nielsen Company. The members of this panel store, in memory, data about the products which they purchase, and forward that data periodically to a central facility. For this purpose, these panelists are generally provided with UPC scanners which they use to scan and store the UPCs attached to the products that they purchase. The scanners are inserted into corresponding docking stations which serve to charge the scanners when the scan-

ners are not in use, and to transmit the UPC information stored in the scanners to the central facility. At the central facility, the data from all panelists are accumulated and correlated in order to generate appropriate reports.

It is also known for a product supplier, such as a retailer, to collect data regarding its product sales so that the product supplier can determine the effectiveness of marketing programs, advertising, promotions, shelf or rack space allocations, product displays, and/or the like. For a retailer, this type of data is generally collected at the point-of-sale terminals where the sales to its customers are processed.

The product supplier might also want to correlate its product sales information with demographic information about its customers so that the product supplier can form conclusions regarding the types of people purchasing its products. For this purpose, it is known for a product supplier to issue customer identification cards which are used by its customers to identify themselves at the time that they make their purchases. Accordingly, the product

supplier can correlate demographic information about its customers with its products sales.

None of the arrangements described above, however, provide the product supplier with information about purchases its customers make from other product suppliers, such as competitive product suppliers. This type of information would also be useful to the product supplier who could then, for example, benchmark the effectiveness of its marketing strategies against those of its competitors.

Therefore, the present invention is directed to a system and method for estimating the purchases which the customers of a product supplier or other marketing entity make from other product suppliers or marketing entities.

#### Summary of the Invention

According to one aspect of the present invention, a method is provided to estimate purchases made by customers of a supplier of interest from other suppliers. The method is performed on a computer and comprises the following steps: a) reading panelist data regarding purchases made by panelists from the supplier of interest and from the other suppliers, wherein the panelists are a subset of the custom-



ers; b) determining a relationship between the purchases made by the panelists from the supplier of interest and the purchases made by the panelists from the other suppliers; c) reading customer data regarding purchases made by the customers from the supplier of interest; and, d) based upon the customer data and the relationship, estimating the purchases made by the customers from the other suppliers.

According to another aspect of the present invention, a method is provided to estimate purchases made by customers of a supplier of interest from other suppliers. The method is performed on a computer and comprises the following steps: a) reading customer data regarding purchases made by the customers from the supplier of interest; b) reading panelist data regarding purchases made by panelists from the supplier of interest and from the other suppliers, wherein the panelists are a subset of the customers; and, c) based upon the customer data and the panelist data, estimating purchases made by the customers from the other suppliers.

According to still another aspect of the present invention, a method is provided to estimate purchases made by customers of a supplier of interest. The method is

performed on a computer and comprises the following steps:  
a) determining a linear relationship between purchases made  
by panelists from the supplier of interest and purchases  
made by the panelists from other suppliers; and, b) esti-  
5 mating purchases by the customers from the other suppliers  
based upon the linear relationship.

According to yet another aspect of the present  
invention, a system for estimating purchases made by custom-  
ers of a supplier of interest comprises analyzing means and  
10 estimating means. The analyzing means analyzes purchases  
made by the customers from the supplier of interest and  
purchases made by panelists from both the supplier of inter-  
est and other suppliers. The panelists are a subset of the  
customers of the supplier of interest. The estimating means  
15 estimates purchases by the customers from the other suppli-  
ers based upon the analyzed purchases.

#### Brief Description of the Drawings

These and other features and advantages of the  
present invention will become more apparent from a detailed  
20 consideration of the invention when taken in conjunction  
with the drawings in which:

Figure 1 is an exemplary computing system which may be used to carry out the present invention; and,

Figures 2A and 2B illustrate a flow chart showing a program that may be executed by the computing system of Figure 1 according to the present invention.

#### Detailed Description

According to the present invention, estimating the purchases made by the customers of a product supplier of interest from other product suppliers is preferably based upon two sources of data. A first source of data provides panelist data. This panelist data is collected from a panel which is formed from the customer base of the product supplier of interest. The panelist data collected from these panelists includes not only data on the purchases that the panelists make from the product supplier of interest, but also data on the purchases that the panelists make from the other product suppliers. Accordingly, the panelist data may be characterized as multichannel data. Also collected from each panelist is a panelist ID uniquely identifying the corresponding panelist.

5 A second source of data is the product supplier of  
interest, who supplies data regarding its sales to its  
customers. This data is referred to hereinafter as customer  
data and may be collected by the product supplier of inter-  
est from its customers using a variety of methods. However,  
a product supplier typically has a preferred customer pro-  
gram in which the product supplier supplies its preferred  
customers with identification cards that can be scanned.  
10 The product supplier then collects purchase data from each  
preferred customer in the usual way and also scans the  
preferred customer's unique identification from the pre-  
ferred customer's identification card. The purchase data  
and the identification data are stored together so that the  
purchase data are associated with the corresponding pre-  
ferred customers. Thus, the data from the second source is  
15 the customer data accumulated from the preferred customers  
of the product supplier of interest and is limited to pur-  
chases made from the product supplier of interest. Accord-  
ingly, the customer data may be characterized as unichannel  
20 data.

Based upon the panelist data and the customer  
data, purchases from other product suppliers made by the

preferred customers of the product supplier of interest can be estimated in accordance with the present invention.

5 This estimate may be made with the aid of a computing system 10 as shown in Figure 1 programmed in accordance with the flow chart shown in Figures 2A and 2B. The computing system 10 includes a data server 12 which stores a data warehouse 14 in its memory. The data warehouse 14, for example, may be a relational database. An access computer 16 is used for writing data into the data warehouse 14 and for reading data out of the data warehouse 14. The access computer 16 may be one or more personal computers and/or workstations. The access computer 16 may be provided with a display in order to present the user of the computing system 10 with various screen displays formatted for ease of data entry, data manipulation, and report generation. The access computer 16 may further have a keyboard, a bar code scanner for scanning codes such as uniform product codes or other industry codes, an optical character recognition device, and/or other data input devices.

20 Data may be supplied to the data warehouse 14 from a variety of sources 18 which include the first and second sources described above. As described below, data may be

arranged in the form of tables such that each table stores a group of similar data. However, it should be understood that, although tables may be referenced herein, the data may be organized in any desired form so that the term "table" as used herein is a generic term.

In order to make the estimate described above, the panelist data are read from the data warehouse 14 at a block 20 of Figures 2A and 2B. The panelist data are aggregated at a block 22 by panelist ID. For each panelist ID, the data is further aggregated by product category. The product categories are selected based on the products offered by the product supplier of interest. For example, if the product supplier of interest is a grocery store chain, one product category might be carbonated beverages and another might be breakfast cereals. Thus, the panelist data are aggregated within each panelist ID table by product category.

Within each product category, the panelist data is further divided between purchases that the corresponding panelist made from the product supplier of interest and purchases that the corresponding panelist made from the other product suppliers. The panelist data are also stored in the product categories by trip. Accordingly, if a panel-

ist purchased a product in a product category from one  
product supplier during one trip and a product in the same  
product category from the same product supplier during  
another trip, the panelist's data would contain an entry in  
the same product category for each trip. Therefore, each  
panelist ID table includes the number of dollars that the  
corresponding panelist spent in each product category by  
trip to the product supplier of interest, and the number of  
dollars that the corresponding panelist spent in each prod-  
uct category by trip to the other product suppliers.

Moreover, the panelist data may be aggregated at  
the block 22 so that the number of dollars spent with the  
product supplier of interest is totaled by panelist ID  
across all product categories and is included in the table  
for the corresponding panelist, and so that the number of  
trips made by each panelist to the product supplier of  
interest is totaled across all product categories and is  
included in the table for the corresponding panelist.  
Similarly, the data may be aggregated at the block 22 so  
that the number of dollars spent with the other product  
suppliers by each panelist is totaled across all product  
categories and is included in the table for the correspond-

ing panelist, and so that the number of trips made by each panelist to the other product suppliers is totaled across all product categories and is included in the table for the corresponding panelist.

5 Furthermore, each panelist ID table further includes a product supplier of interest share for each product category. In determining this share, the dollars paid by the panelist to the product supplier of interest in the corresponding product category during all trips covered by the applicable time period are divided by the total dollars paid by the panelist to the product supplier of interest in all product categories during the same time period. Each panelist ID table includes the dollars paid by the panelist to other product suppliers for each product category during 10 all trips covered by the applicable time period.

15 Finally, the panelist data in each panelist ID table are also similarly aggregated for each department. That is, the panelist data are aggregated in the same way as discussed above but this time by department, where each 20 department covers one or more related product categories. Thus, the panelist data within a department are aggregated



according to trips, dollars, share, etc. for the product supplier of interest and for the other product suppliers.

It should be understood from the foregoing description that the panelist data stored in the panelist ID tables are segregated between product supplier of interest  
5 and other product suppliers.

At a block 24, the customer data (from the product supplier of interest, i.e., the second source) are read from the data warehouse 14. At a block 26, the customer data are  
10 aggregated by customer ID in the same manner as described above in connection with the block 22.

At a block 28, an unrotated principal components factor analysis is performed on the data aggregated at the block 26 (i.e., the aggregated customer data). Thus, except  
15 for inconsistent customer data, the inputs to the unrotated principal components factor analysis are all records stored in the customer ID tables, where each record contains the purchases at the UPC level that the corresponding customer made in a specific store at a specific date and time. If a  
20 customer does not spend money in any product category in at least two trips out of the most recent sixteen week period, then the data for that customer is considered to be incon-

5       sistent customer data. Also, if a customer does not spend  
at least a predetermined amount of money (such as \$75.00) in  
ant product category during a predetermined amount of time  
(such as a year), the data for that customer is considered  
to be inconsistent customer data. Furthermore, if customer  
data is collected for a product that does not fit into a  
defined product category, such customer data is considered  
to be inconsistent customer data. Thus, inconsistent cus-  
tomer data is not used by the block 28.

10       The unrotated principal components factor analysis  
is a well-known statistical analytical tool for analyzing  
input data. This analysis produces a factor matrix which is  
a  $k \times i$  matrix having  $k$  rows and  $i$  columns, where  $k$  is the  
number of customer IDs, and where  $i$  is the number of factors  
15       resulting from the unrotated principal components factor  
analysis. The unrotated principal components factor analy-  
sis collapses the  $j$  dimensions in product category space  
(where  $j$  is the number of product categories) down to  $i$   
dimensions. The value of  $i$  may be selected so that each of  
20       the  $i$  dimensions has a minimum eigenvalue (such as 1.3).

At a block 30, the factor matrix generated at the  
block 28 is used to score the panelist data. This scoring

is accomplished by matrix multiplying the factor matrix produced at the block 28 and the  $k \times j$  panelist data to produce  $k \times i$  factors. Accordingly, this matrix multiplication generates a panelist set of factors  $F_1$  through  $F_i$  for each panelist ID, where  $i \leq j$ . Thus, a principal component category may be identical to a product category if the sales in the product category are sufficiently high. The factors  $F_1$  through  $F_i$  of the panelist set of factors are part of the panelist predictor variables discussed below.

Similarly, at a block 32, the factor matrix generated at the block 28 is also used to score the customer data. Again, this scoring is accomplished by matrix multiplying the factor matrix produced at the block 28 and the  $k \times j$  customer data to produce  $k \times i$  factors. This matrix multiplication, therefore, generates a customer set of factors  $F_1$  through  $F_i$  for each customer ID. The factors  $F_1$  through  $F_i$  of the customer set of factors are part of the customer predictor variables discussed below.

For each panelist, other panelist predictor variables are created at a block 34. These other panelist variables are determined from the panelist data and include one or more of the following:  $F_1^2$  through  $F_i^2$  which are the

squares of the corresponding factors  $F_1$  through  $F_i$  created at the block 30; interdependent factors which include the products of all possible pairs of the factors  $F_1$  through  $F_i$  created at the block 30 (that is,  $F_1 \times F_2$ ,  $F_1 \times F_3$ , . . .  $F_1 \times F_i$ ,  $F_2 \times F_3$ ,  $F_2 \times F_4$ , . . .  $F_2 \times F_i$ ,  $F_3 \times F_4$ , . . .  $F_{i-1} \times F_i$ );  $T_1$ ,  $T_2$ , . . . ,  $T_j$  which are the total number of panelist trips in the corresponding  $j$  product categories;  $T_1^2$ ,  $T_2^2$ , . . . ,  $T_j^2$  which are the squares of  $T_1$ ,  $T_2$ , . . . ,  $T_j$ ;  $TD$  which is the sum of the dollars spent by the corresponding panelist in all product categories;  $TD^2$  which is the square of  $TD$ ;  $CD_1$ ,  $CD_2$ , . . . ,  $CD_j$  which are the dollars spent by the corresponding panelist with the product supplier of interest in the corresponding  $j$  product categories; and,  $CO_1$ ,  $CO_2$ , . . . ,  $CO_j$  which are the dollars spent with the other product suppliers in the corresponding  $j$  product categories.

Similarly, for each customer ID, other customer predictor variables are created at a block 36 for each product category. These customer predictor variables are determined from the customer data and include one or more of the following:  $F_1^2$  through  $F_i^2$  which are the squares of the corresponding factors  $F_1$  through  $F_i$  created at the block 32; interdependent factors which include the products of all

possible pairs of the factors  $F_1$  through  $F_i$  created at the  
block 32 (that is,  $F_1 \times F_2, F_1 \times F_3, \dots, F_1 \times F_i, F_2 \times F_3,$   
 $F_2 \times F_4, \dots, F_2 \times F_i, F_3 \times F_4, \dots, F_{i-1} \times F_i$ );  $T_1, T_2,$   
 $\dots, T_j$  which are the total number of customer trips in the  
corresponding  $j$  product categories;  $T_1^2, T_2^2, \dots, T_j^2$  which  
are the squares of  $T_1, T_2, \dots, T_j$ ;  $TD$  which is the sum of  
the dollars spent by the corresponding customer in all  
product categories;  $TD^2$  which is the square of  $TD$ ; and,  
 $CD_1, CD_2, \dots, CD_j$  which are the dollars spent by the corre-  
sponding customer with the product supplier of interest in  
the corresponding  $j$  product categories.  $CO_1, CO_2, \dots, CO_j,$   
which are the dollars spent by the corresponding customer  
with the other product suppliers in the corresponding prod-  
uct categories, is to be estimated.

A set of criterion variables  $CV_i$  is created at a  
block 38 for each product category by dividing the panelist  
IDs into buckets according to their values of  $CO_j$ . That is,  
for a first product category, a bucket zero contains all  
panelist IDs whose corresponding value of  $CO_1 = 0$ , provided  
that there are at least a predetermined number (such as 150)  
of such panelist IDs. The remaining panelist IDs are sorted  
from highest to lowest according to their values of  $CO_1$  and

are then divided evenly into buckets one through n for the first product category, with the bucket one containing the remaining panelist IDs whose values of  $CO_1$  are lowest (other than zero), with the bucket two containing the remaining panelist IDs whose values of  $CO_1$  are next lowest, and so on. Each of the buckets one through n must contain at least the predetermined number of panelist IDs, and each of the buckets one through n must come as close as possible to containing the predetermined number of panelist IDs with the proviso that all of the buckets one through n should contain, as closely as possible, an equal number of panelist IDs.

If the bucket zero is only a few panelist IDs short of the predetermined number, a sufficient number of remaining panelist IDs whose values of  $CO_1$  are lowest may be moved into the bucket zero so that the bucket zero contains the predetermined number of panelist IDs. This movement is made before the sorting and dividing described above. On the other hand, if the bucket zero contains only a few panelist IDs, no panelist IDs are put into the bucket zero and instead all panelist IDs are sorted and divided as described above. The criteria used to make these decisions may be established as desired.

The criterion variable  $CV_0$  for the first product category is set equal to the number of panelist IDs in the bucket zero, the criterion variable  $CV_1$  for the first product category is set equal to the number of panelist IDs in the bucket one, the criterion variable  $CV_2$  for the first product category is set equal to the number of panelist IDs in the bucket two, and so on.

This process is then repeated for each of the other product categories so that there is a set of criterion variables for each of the product categories.

At a block 40, the panelist data are split between model data and leave out data. All panelist data associated with a randomly selected  $r\%$  of the panelist IDs are designated as leave out data. The value of  $r$ , for example, may be 15, so that all panelist data associated with 15% of the panelist IDs selected at random are designated as leave out data. The remaining panelist data is designated as model data.

At a block 42, scoring rules are determined from the model data. For example, a commercially available software program entitled "Wizwhy" supplied by WizSoft can be used at the block 42 in order to determine these scoring

rules. These scoring rules are if-then scoring rules and are subsequently used in creating new panelist and customer predictor variables as described below. Wizwhy actually creates two kinds of if-then scoring rules, i.e., "is if-then" scoring rules and "is not if-then" scoring rules. An "is if-then" scoring rule, for example, has the following format: if dollars are greater than 100, then the output is true (1), otherwise the output is false (0). Similarly, an "is not if-then" scoring rule, for example, has the following format: if dollars are less than 50, then the output is false (0), otherwise the output is true (1). The inputs to the Wizwhy program during this iteration are the predictor variables which meet all of the following three criteria: (i) the predictor variables must correspond to the model data; (ii) the predictor variables must correspond to the panelist IDs in the bucket zero; and, (iii) the predictor variables must correspond to the first product category. However, the predictor variables used at the block 42 exclude the square terms and the interdependent factors.

Next, the Wizwhy program determines similar scoring rules based on the panelist IDs in the bucket one. That is, the inputs to the Wizwhy program during this iteration



are the predictor variables which meet all of the following three criteria: (i) the predictor variables must correspond to the model data; (ii) the predictor variables must correspond to the panelist IDs in the bucket one; and, (iii) the predictor variables must correspond to the first product category. However, the predictor variables used at the block 42 again exclude the square terms and the interdependent factors. An iteration of the Wizwhy program is similarly executed for each of the other buckets. This process is then repeated for each of the other product categories.

At a block 44, new panelist predictor variables are created based upon both the model data and the leave out data. The creation of these new panelist predictor variables is effected by using first the "is if-then" scoring rules and then the "is not if-then" scoring rules generated at the block 42 in order to score the model data and the leave out data by bucket and by product category. That is, the model data and the leave out data which are in a first of the product categories and which relate to the panelist IDs in the bucket zero are first supplied to the "is if-then" scoring rules and the outputs (0's and 1's) are summed. This sum is a first new panelist predictor variable

relating to data which are in the first product category and which correspond to the panelist IDs in the bucket zero.

5           Next, the model data and the leave out data which are in this first product category and which correspond to the panelist IDs in the bucket one are supplied to the "is if-then" scoring rules and the outputs (0's and 1's) are summed. This sum is a first new panelist predictor variable relating to data which are in the first product category and which correspond to the panelist IDs in the bucket one.  
10       This process, using the "is if-then" scoring rules, is repeated so that there is a first new panelist predictor variable for the first product category relating to each bucket.

15           Then, the model data and the leave out data which are in the first product category and which relate to the panelist IDs in the bucket zero are supplied to the "is not if-then" scoring rules and the outputs (0's and 1's) are summed. This sum is a second new panelist predictor variable relating to data which are in the first product category and which correspond to the panelist IDs in the bucket  
20       zero.

5           Next, the model data and the leave out data which  
are in this first product category and which correspond to  
the panelist IDs in the bucket one are supplied to the "is  
not if-then" scoring rules and the outputs (0's and 1's) are  
summed. This sum is a second new panelist predictor vari-  
10       able relating to data which are in the first product cate-  
gory and which correspond to the panelist IDs in the bucket  
one. This process, using the "is not if-then" scoring  
rules, is repeated so that there is a second new panelist  
predictor variable for the first product category relating  
to each bucket.

15           Then, the model data and the leave out data which  
are in a second of the product categories and which relate  
to the panelist IDs in the bucket zero are supplied to the  
"is if-then" scoring rules and the outputs (0's and 1's) are  
summed. This sum is a first new panelist predictor variable  
relating to data which are in the second product category  
and which correspond to the panelist IDs in the bucket zero.

20           Next, the model data and the leave out data which  
in this second product category and which correspond to the  
panelist IDs in the bucket one are supplied to the "is if-  
then" scoring rules and the outputs (0's and 1's) are

summed. This sum is a first new panelist predictor variable relating to data which are in the second product category and which correspond to the panelist IDs in the bucket one. This process, using the "is if-then" scoring rules, is repeated so that there is a first new panelist predictor variable for the second product category relating to each bucket.

Then, the model data and the leave out data which are in the second product category and which relate to the panelist IDs in the bucket zero are first supplied to the "is not if-then" scoring rules and the outputs (0's and 1's) are summed. This sum is a second new panelist predictor variable relating to data which are in the second product category and which correspond to the panelist IDs in the bucket zero.

Next, the model data and the leave out data which in this second product category and which correspond to the panelist IDs in the bucket one are supplied to the "is not if-then" scoring rules and the outputs (0's and 1's) are summed. This sum is a second new panelist predictor variable relating to data which are in the second product category and which correspond to the panelist IDs in the bucket

one. This process, using the "is not if-then" scoring rules, is repeated so that there is a second new panelist predictor variable for the second product category relating to each bucket.

5           This process is repeated for each of the other product categories.

10           At a block 46, new customer predictor variables are created based upon the customer data in a similar manner. However, in this case, there are no buckets. Thus, the creation of two new customer predictor variables per product category is effected by using first the "is if-then" scoring rules and then the "is not if-then" scoring rules generated at the block 42 in order to score the customer data for each corresponding product category. That is, the  
15           customer data in the first product category are first supplied to the "is if-then" scoring rules and the outputs (0's and 1's) are summed. This sum is the first new customer predictor variable for the customer data relative to the first product category. Then, the customer data in the  
20           first product category are supplied to the "is not if-then" scoring rules and the outputs (0's and 1's) are summed. This sum is the second new customer predictor variable for

the customer data relative to the first product category. This process is repeated for each of the other product categories so that there are two new customer predictor variables based on customer data for each product category.

At a block 48, the subroutine Proc Reg of the commercially available program SAS is performed. The Proc Reg subroutine is a linear regression that is performed based only on the model data and that generates an output coefficient matrix. Each row of this matrix contains a set of coefficients for a corresponding product category. The Proc Reg subroutine is performed J times, where J is 50 for example, according to the following equation:

$$\text{model depvar} = \text{indepvar} / \text{maxr} \quad \text{stop} = i \quad (1)$$

where the dependent variable depvar are the criterion variables by product category as determined at the block 38, and where the independent variables indepvar are the predictor variables created at the blocks 44 by product category. The Proc Reg subroutine is shown in the attached Appendix. The output coefficient matrix of the Proc Reg subroutine as executed by the block 48 is a linear equation for each

product category, where each linear equation has a set of coefficients as contained in a corresponding row of the output coefficient matrix. These linear equations are used by the block 50 as discussed below. These linear equations establish a linear relationship by product category between purchases made by panelists from the supplier of interest and purchases made by the panelists from the other suppliers.

The block 50 uses the new customer predictor variables determined at the block 46 as inputs for the variables in the linear equations determined at the block 48 in order to estimate (i.e., forecast) sales to the preferred customers by the other product suppliers in each of the product categories. Accordingly, the new predictor variables, which are created at the block 46 from the customer data and which relate to the purchases made by the customers from the supplier of interest in each product category, are substituted for the variables in the corresponding linear equation, which relates the purchases made by the panelists from the supplier of interest to the purchases made by the panelists from the other suppliers, in order to forecast the purchases that the customers made from the other suppliers.

As used herein, a product supplier, for example, may be a retailer, a wholesaler, a manufacturer, or other product supplier. Also, although the present invention has been described in detail in terms of products for ease of understanding, it should be understood that the present invention applies equally well to services. Therefore, the present invention should be understood to apply to products and/or services.

Certain modifications of the present invention have been discussed above. Other modifications will occur to those practicing in the art of the present invention. For example, as described above, the present invention is implemented by a program represented by the flow chart shown in Figures 2A and 2B. Instead, the present invention may be implemented by way of a neural network arranged in accordance with the processing represented by Figures 2A and 2B.

Also, as described above, customer data may be collected from the preferred customers of the product supplier of interest. However, it should be understood that the customer data may be collected from more, fewer, and/or different customers of the product supplier of interest.



Therefore, the term "preferred customer" is used generically herein.

Moreover, certain predictor variables as described above are created at the blocks 30-36. However, fewer  
5 and/or different predictor variables may instead be created.

Furthermore, as described above, if a customer does not spend money in any product category in at least two trips out of the most recent sixteen week period, then the data for that customer is considered to be inconsistent  
10 customer data. However, the number of trips may be other than two, and/or the time period may be other than a sixteen week period.

Accordingly, the description of the present invention is to be construed as illustrative only and is for the purpose of teaching those skilled in the art the best mode  
15 of carrying out the invention. The details may be varied substantially without departing from the spirit of the invention, and the exclusive use of all modifications which are within the scope of the appended claims is reserved.

WHAT IS CLAIMED IS:

1                   1. A method of estimating purchases made by  
2 customers of a supplier of interest from other suppliers,  
3 wherein the method is performed on a computer, and wherein  
4 the method comprises the following steps:

5                   a) reading panelist data regarding purchases made  
6 by panelists from the supplier of interest and from the  
7 other suppliers, wherein the panelists are a subset of the  
8 customers;

9                   b) determining a relationship between the pur-  
10 chases made by the panelists from the supplier of interest  
11 and the purchases made by the panelists from the other  
12 suppliers;

13                   c) reading customer data regarding purchases made  
14 by the customers from the supplier of interest; and,

15                   d) based upon the customer data and the relation-  
16 ship, estimating the purchases made by the customers from  
17 the other suppliers.

1                   2. The method of claim 1 wherein step a) com-  
2                   prises the step of aggregating the panelist data according  
3                   to categories, and wherein step (c) comprises the step of  
4                   aggregating the customer data according to categories.

1                   3. The method of claim 2 wherein the aggregated  
2                   data includes a number of dollars that each panelist spent  
3                   with the supplier of interest by category and a number of  
4                   dollars that each panelist spent with the other suppliers by  
5                   category.

1                   4. The method of claim 3 wherein the aggregated  
2                   data includes share for the supplier of interest and share  
3                   for the other suppliers by category.

1                   5. The method of claim 2 comprising the step of  
2                   performing an unrotated principal components factor analysis  
3                   on at least one of the aggregated panelist data and the  
4                   aggregated customer data.

1           6. The method of claim 5 comprising the step of  
2 determining predictor variables based upon on at least one  
3 of the aggregated panelist data and the aggregated customer  
4 data.

1           7. The method of claim 6 wherein the predictor  
2 variables include the following: factors  $F_1$  through  $F_i$   
3 resulting from the performing step; a total number of trips  
4 in which dollars were spent in a category; and, a total  
5 number of dollars spent in a category.

1           8. The method of claim 7 wherein the predictor  
2 variables also include the following: the squares of the  
3 factors  $F_1$  through  $F_i$ ; interdependent factors based upon  
4 products of the factors  $F_1$  through  $F_i$ ; a square of the  
5 total number of trips; and, a square of the total number of  
6 dollars.

1           9. The method of claim 6 wherein the predictor  
2 variables include factors  $F_1$  through  $F_i$  resulting from the  
3 performing step.

1                   10. The method of claim 9 wherein the predictor  
2 variables also include the squares of the factors  $F_1$  through  
3  $F_i$ .

1                   11. The method of claim 9 wherein the predictor  
2 variables also include interdependent factors based upon  
3 products of the factors  $F_1$  through  $F_i$ .

1                   12. The method of claim 6 wherein the predictor  
2 variables include a total number of trips in which dollars  
3 were spent in a category.

1                   13. The method of claim 12 wherein the predictor  
2 variables also include a square of the total number of  
3 trips.

1                   14. The method of claim 6 wherein the predictor  
2 variables include a total number of dollars spent in a  
3 category.

1           15. The method of claim 14 wherein the predictor  
2 variables also include a square of the total number of dollars.

1           16. The method of claim 6 comprising the step of  
2 determining criterion variables based upon at least one of  
3 the aggregated panelist data and the aggregated customer  
4 data.

1           17. The method of claim 16 wherein the step of  
2 determining criterion variables comprises the step of divid-  
3 ing the panelists into buckets and of determining the crite-  
4 rion variables as the number of panelists in each bucket.

1           18. The method of claim 16 comprising the steps  
2 of executing a routine in order to generate a set of scoring  
3 rules, and creating new predictor variables based upon the  
4 scoring rules.

1           19. The method of claim 18 wherein the step of  
2 creating new predictor variables based upon the scoring  
3 rules comprises the steps of inputting the panelist data and  
4 the customer data to the scoring rules by product category

5 and by bucket of panelist IDs and summing an output of the  
6 scoring rules by product category and by bucket.

1 20. The method of claim 18 comprising the step of  
2 performing a linear regression based upon the new predictor  
3 variables and the criterion variables in order to generate  
4 the relationship, wherein the relationship is a linear  
5 relationship.

1 21. The method of claim 20 wherein step d) com-  
2 prises the step of applying the customer data to the linear  
3 relationship.

1 22. The method of claim 1 comprising the step of  
2 performing an unrotated principal components factor analysis  
3 on at least one of the panelist data and the customer data.

1 23. The method of claim 22 comprising the step of  
2 determining predictor variables based upon on at least one  
3 of the panelist data and the customer data.

1                   24. The method of claim 23 comprising the step of  
2 determining criterion variables based upon on at least one  
3 of the panelist data and the customer data.

1                   25. The method of claim 24 comprising the step of  
2 performing a linear regression based upon the predictor  
3 variables and the criterion variables in order to generate  
4 the relationship, wherein the relationship is a linear  
5 relationship.

1                   26. The method of claim 25 wherein step d) com-  
2 prises the step of applying the customer data to the linear  
3 relationship in order to estimate the purchases made by the  
4 customers from the other suppliers.

1                   27. The method of claim 1 comprising the step of  
2 performing an unrotated principal components factor analysis  
3 on the customer data.



1           28. The method of claim 27 comprising the step of  
2 performing a linear regression based upon the panelist data  
3 in order to generate the relationship, wherein the relation-  
4 ship is a linear relationship.

1           29. The method of claim 28 wherein step d) com-  
2 prises the step of applying the customer data to the linear  
3 relationship.

1           30. A method of estimating purchases made by  
2 customers of a supplier of interest from other suppliers,  
3 wherein the method is performed on a computer, and wherein  
4 the method comprises the following steps:

5           a) reading customer data regarding purchases made  
6 by the customers from the supplier of interest;

7           b) reading panelist data regarding purchases made  
8 by panelists from the supplier of interest and from the  
9 other suppliers, wherein the panelists are a subset of the  
10 customers; and,

11           c) based upon the customer data and the panelist  
12 data, estimating purchases made by the customers from the  
13 other suppliers.

1                   31. The method of claim 30 wherein step c) com-  
2                   prises the step of aggregating the customer data and the  
3                   panelist data according to categories.

1                   32. The method of claim 31 wherein step c) com-  
2                   prises the step of performing an unrotated principal compo-  
3                   nents factor analysis on at least a portion of the aggre-  
4                   gated data.

1                   33. The method of claim 32 wherein step c) com-  
2                   prises the step of determining predictor variables based  
3                   upon the performing step and upon at least a portion of the  
4                   aggregated data.

1                   34. The method of claim 33 wherein step c) com-  
2                   prises the step of performing a linear regression on the  
3                   predictor variables in order to generate a linear equation  
4                   for each category.

1                   35. The method of claim 34 wherein step c) com-  
2                   prises the step of estimating the purchases made by the  
3                   customers from the other suppliers in each category by  
4                   plugging the customer data into the linear equation for each  
5                   category.

1                   36. The method of claim 30 wherein step c) com-  
2                   prises the step of performing an unrotated principal compo-  
3                   nents factor analysis based upon at least one of the panel-  
4                   ist data and the customer data.

1                   37. The method of claim 36 wherein step c) com-  
2                   prises the step of creating a linear equation based upon  
3                   results from the unrotated principal components factor  
4                   analysis.

1                   38. The method of claim 37 wherein step c) com-  
2                   prises the step of estimating the purchases made by the  
3                   customers from the other suppliers by plugging the customer  
4                   data into the linear equation.

1                   39. A method of estimating purchases made by  
2 customers of a supplier of interest, wherein the method is  
3 performed on a computer, and wherein the method comprises  
4 the following steps:

5                   a) determining a linear relationship between  
6 purchases made by panelists from the supplier of interest  
7 and purchases made by the panelists from the other suppli-  
8 ers; and,

9                   b) estimating purchases by the customers from the  
10 other suppliers based upon the linear relationship.

1                   40. The method of claim 39 wherein step b) com-  
2 prises the step of estimating purchases from the other  
3 suppliers made by the customers of the supplier of interest  
4 based upon the linear relationship and purchases made by the  
5 customers from the supplier of interest.

1                   41. The method of claim 39 wherein the panelists  
2 are a subset of the customers.

1           42. The method of claim 41 wherein step b) com-  
2       prises the step of estimating purchases from the other  
3       suppliers made by the customers of the supplier of interest  
4       based upon the linear relationship and purchases made by the  
5       customers from the supplier of interest.

1           43. A system for estimating purchases made by  
2       customers of a supplier of interest comprising:

3           analyzing means for analyzing purchases made by  
4       the customers from the supplier of interest and purchases  
5       made by panelists from both the supplier of interest and  
6       other suppliers, wherein the panelists are a subset of the  
7       customers of the supplier of interest; and,

8           estimating means for estimating purchases by the  
9       customers from the other suppliers based upon the analyzed  
10      purchases.

1           44. The method of claim 43 wherein the analyzing  
2       means comprises means for performing an unrotated principal  
3       components factor analysis based upon purchase data.

1                   45. The method of claim 44 wherein the analyzing  
2 means comprises means for determining a linear relationship  
3 based upon results from the unrotated principal components  
4 factor analysis.

1                   46. The method of claim 45 wherein the linear  
2 relationship relates purchasers made by the panelists from  
3 the supplier of interest to purchases made by the panelists  
4 from the other suppliers.

1                   47. The method of claim 45 wherein the estimating  
2 means estimates the purchases by the customers from the  
3 other suppliers based upon the purchases by the customers  
4 from the supplier of interest and upon the linear relation-  
5 ship.

1                   48. The method of claim 43 wherein the analyzing  
2 means comprises means for determining a linear relationship  
3 between purchasers made by the panelists from the supplier  
4 of interest and purchases made by the panelists from the  
5 other suppliers.

1                   49. The method of claim 48 wherein the estimating  
2 means estimates the purchases by the customers from the  
3 other suppliers by plugging the purchases by the customers  
4 from the supplier of interest into upon the linear relation-  
5 ship.

ABSTRACT OF THE DISCLOSURE

Purchases made by customers of a supplier of interest from other suppliers are estimated based upon panelist data regarding purchases made by panelists from the supplier of interest and from the other suppliers, and based upon customer data regarding purchases made by the customers from the supplier of interest. The panelists are a subset of the customers. A linear relationship between the purchases made by the panelists from the supplier of interest and the purchases made by the panelists from the other suppliers is determined. Based upon the customer data and the linear relationship, the purchases made by the customers from the other suppliers are estimated.

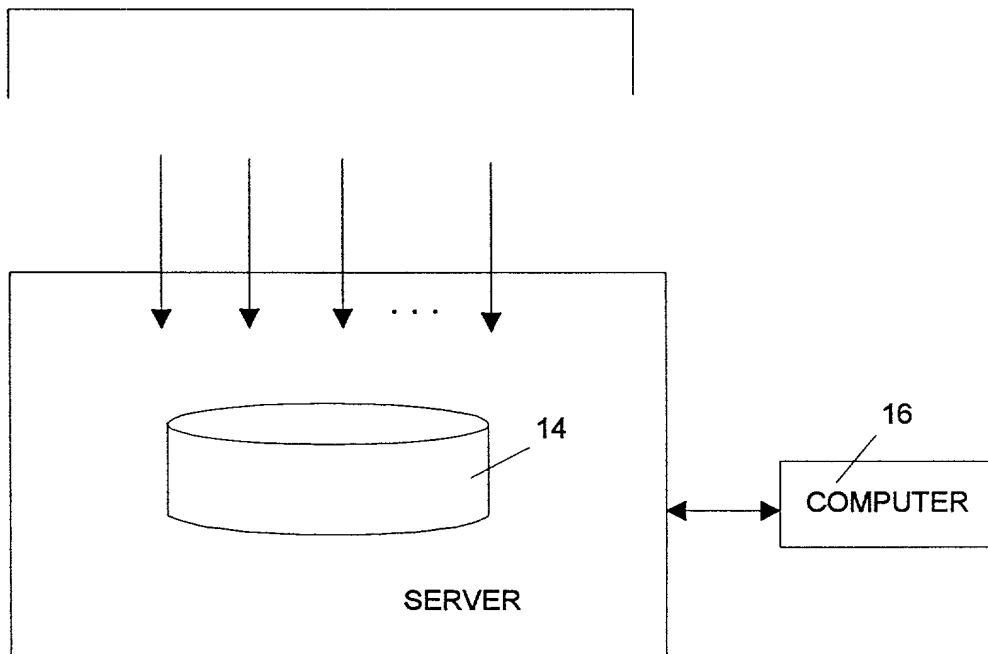


## APPENDIX

```
> proc reg data=abase outest=outest outsscp=outsscp noprint ;  
>   model aa_dpvar = &ivar / selection=maxr stop=&i ;  
> run ;  
>
```

DATA=ABASE OUTEST=OUTEST OUTSSCP=OUTSSCP NOPRINT ;  
MODEL AA\_DPVAR = &IVAR / SELECTION=MAXR STOP=&I ;  
RUN ;  
>

18



12

10

FIGURE 1

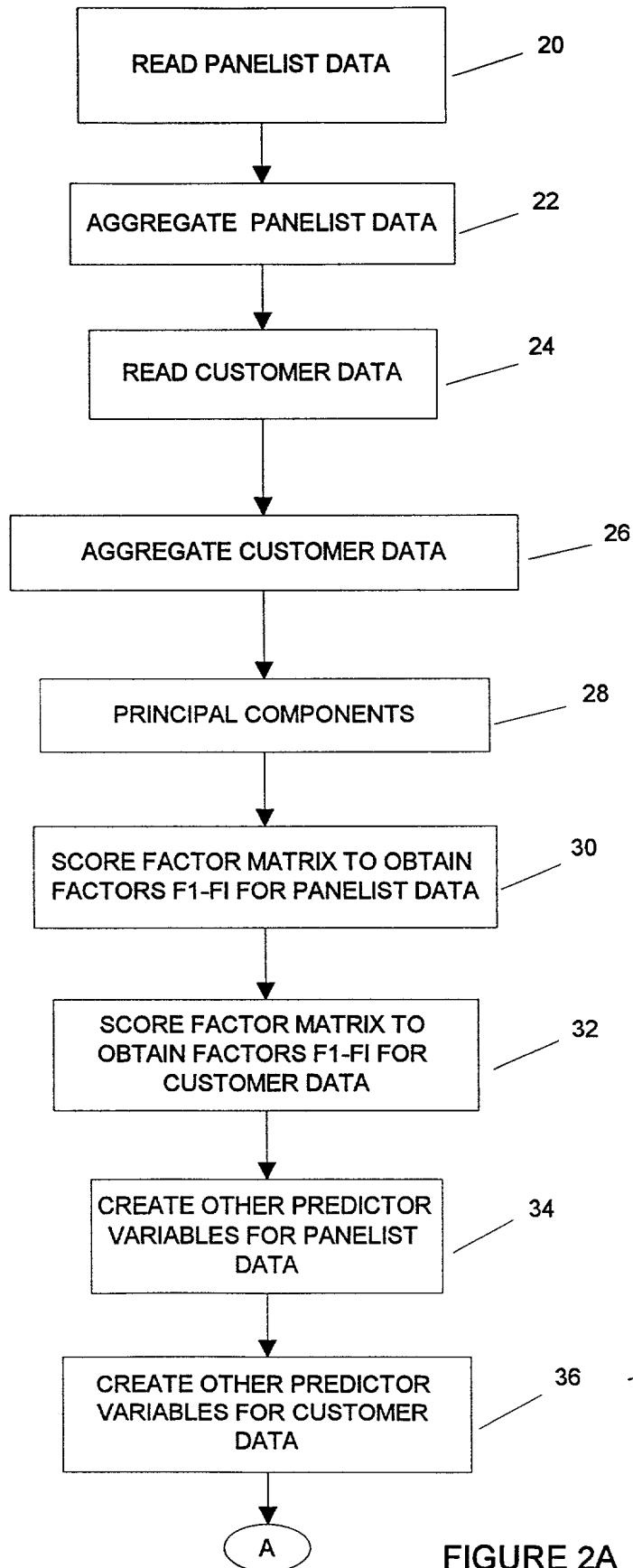


FIGURE 2A

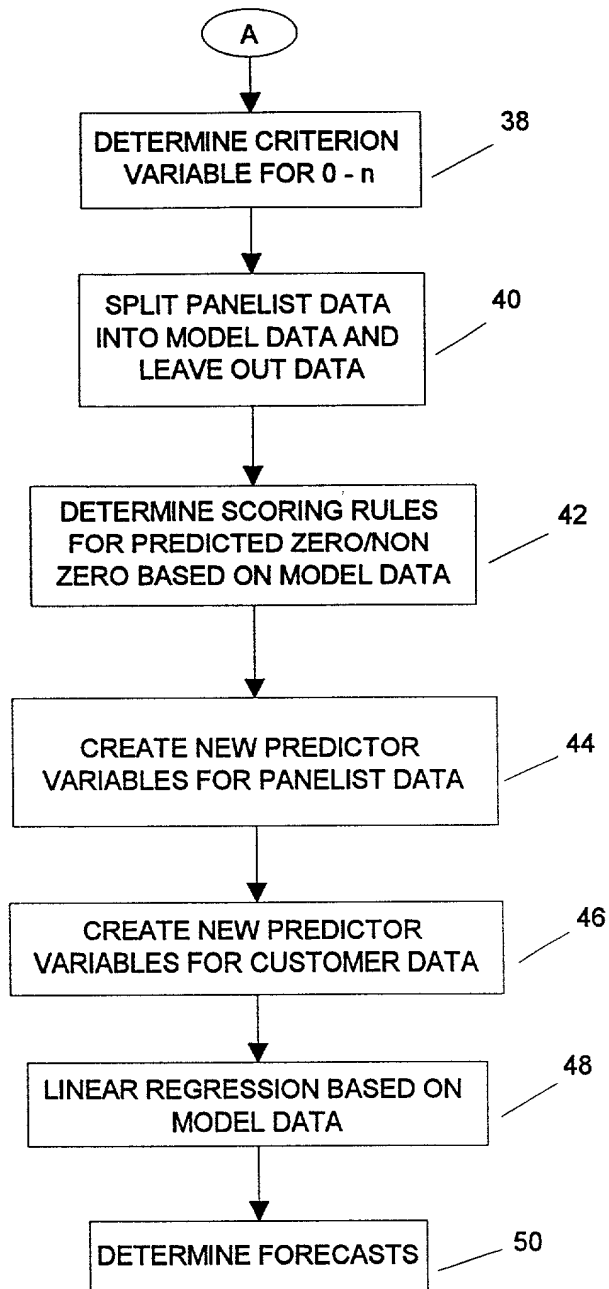


FIGURE 2B